Forensic analysis of IoT ecosystem

François Bouchaud, Thomas Vantroys, Gilles Grimaud

Forensic analysis of IoT ecosystem

François Bouchaud
C3N - National cyber-crime unit
Gendarmerie Nationale
francois.bouchaud@gendarmerie.interieur.gouv.fr

Thomas Vantroys, Gilles Grimaud
Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL
F-59000 Lille, France
Univ. Lille, CNRS, USR 3380 - IRCICA F-59000 Lille, France
{thomas.vantroys,gilles.grimaud}@univ-lille.fr

***Abstract*—Connected toys, home automation appliances and wearable devices are all part of the Internet of Things (IoT). They digitize our lives and generate massive data transfer. This phenomenon brings many opportunities for police investigations, with a rich source of evidence[1], information that is often unexplored and not valued. The proliferation of connected devices has given rise to IoT Forensic. It deals with IoT-related cyber-crime. It includes both local and external investigations. It is therefore a much more complex, multidimensional, and multidisciplinary approach than traditional forensics. Data can be scattered throughout the infrastructure, depending on its management policy (synchronization and storage). They become meaningful when they are contextualized and cross-referenced. This article focus on the process of exploiting data from connected devices at a crime scene and on reconstructing the timeline of a criminal event. We use a scenario based on real facts to illustrate the process of IoT investigation. Based on discovered data and objects, we reconstruct the chronology of criminal events.**

***Keywords*-Internet of Things, IoT Forensic, Investigations, Crime scene, Collection and Analysis, Evidence.**

I. Introduction

More and more objects around us are becoming digital and accessing the Internet. The term "Internet of Things"(IoT) is used to refer to this structuring of the connected infrastructure. In its early days, it referred to the networking of objects equipped with radio-frequency chips (RFID). The concept was then democratized and generalized with the rise of wireless networks, the cloud and the miniaturization of embedded systems. The literature defines it as a 'group of infrastructure interconnecting connected objects and allowing their management, data mining and the access to data they generate' [1]. This definition retains a certain neutrality of use. It integrates technological identification, exploitation context and issues related to the study of polymorphic information. It therefore refers to a notion of extended architecture, structuring an ecosystem of connected devices in order to offer new services.

The spread of this all-digital phenomenon is accelerating. With the exception of the most popular of these, such as connected watches or home automation, they are barely visible, but they are constantly scrutinizing and questioning our daily lives. This phenomenon generates massive data transfers in the information system (IS). This environment constitutes an unprecedented receptacle of data, a real opportunity for police investigations in the search for the truth. This is the case, for example, of the connected thermometer *Nest* with learning capabilities. Coupled with the home ecosystem, it is capable of triggering actions such as turning on the heating when the phone is recognized in a near field. This information can be used to reconstruct events that have occurred and to determine presence in a house. This reading and understanding of the connected environment brings new challenges for investigators. The analysis phase is even more complex when data is dispersed and/or fragmented within the connected infrastructure, both locally and online. This determination of the presence and positioning of information remains specific to each environment. In addition to this problem, there are dependencies within ecosystems through 'hidden links'. The same result has several causes. How to approach the analysis phase when faced with this scattering of data? How to associate the right data to the right equipment? What are the criteria for event reconstruction? How to establish the dependencies between the equipment? Which actors caused an event? What is the data path in the connected infrastructure? What is the meaning of the result of the traces analysis?

This article proposes an analytical framework for leveraging data from connected devices at a crime scene to determine the timeline of a criminal event. We use a real-world based crime scenario to illustrate our approach.

Section II of this article covers previous work in the field of forensics science in IoT environments; Section III describes the forensic analysis of devices from a crime scene; Section IV discusses the contextualization of the data and the reconstruction of the chronology of events; Section V provides the conclusion of the paper and the next stage of this research.

II. Related Works

Evidence from IoT can be retrieved from household appliances, cars, RFID readers, etc. The sources therefore differ in nature, number, format and protocols used. Data is fragmented and dispersed. They become a whole in the global architecture. To understand it, the investigator relies on the practices of Digital Forensics and Cloud Forensics.

A. Forensic Artefacts from IoT Products

The scientific literature is rich in works on the study of connected objects and their artefacts, whether they are wearable devices [2]–[5] or voice assistants [6]–[8] and more globally any connected device of daily life [9]–[11]. These devices are characterized by their own data formats, protocols and

[1]In this paper, the term 'evidence' must be understood in legal and forensic sense.

physical interfaces [12]. Some articles focus on the analysis of data generated when using mobile applications [13], [14]. In particular, they refer to synchronized data, SQLite databases and cache files containing connection information to IoT platforms. There is also work on understanding event logs [15], [16] and analyzing local network flow [17]–[19], particularly in the context of intrusion detection [20], [21].

B. Cloud Forensics

The cloud contains rich sources of evidence due to its role as an information hub. Much research has focused on describing the challenges of conducting digital investigations in the cloud [22]–[24]. The studies offer two perspectives: client-based cloud forensics and cloud-native forensics. The client-based approach involves acquiring and analyzing data recorded locally by applications or web browsers in relation to the use of cloud services. With the development of new cloud service platforms, research has also focused on cloud-native forensics. The challenge is to deal with large amounts of data that are not stored in traditional devices or simply in temporary caches.

Another challenge is the issue of jurisdictional boundaries and the lack of third-party agreement [25]. Data can transit between other devices or IoT services in the cloud. Evidence collection from the cloud is another drawback, as is its physical inaccessibility. Some interesting work proposes models or solutions that could address the inherent problems of preserving digital evidence and its integrity in cloud computing [26], [27].

C. Limits of a Traditional and Unitary Approach

Most studies remain flawed by focusing on a specific object or on a local architecture composed of devices of the same family, such as a connected home and its home automation system. These approaches omit the compatibility of inter-object connections and the new dependencies between systems as well as the dispersion of information in the infrastructure according to the configurations and services offered. A device is likely to be controlled or accessed from hardware separate from the system, according to a hidden link structure. Thus, data is propagated and stored in network devices, contributing or not to the target object [28]. However, the IoT is increasingly mixing connected devices from different families from versatile modules. Ecosystems are customized based on user choices and configurations. For example, voice assistants and their native voice command solution link connected objects in the same home, whether they are home automation, appliances or security devices. Initially classified as simple connected devices, they have become true advanced ecosystems with intelligence. This common service has recently been deployed in objects outside the home through connected bracelets or complex systems such as a connected car. In the same way that the multimedia of a connected vehicle contains the data of a synchronized phone, this highly constrained in-vehicle system is likely to contain the information of the home or an individual, and vice versa. The application layer is the binder for the exchange of useful information between different hardware environments, mixing data from all horizons. The boundaries between connected systems are becoming increasingly porous. The market tends to evolve in this direction given to the development of communication protocols around interoperability and the creation of partnerships between companies in the sector, particularly through online platforms or shared solutions such as virtual personal assistants. This evolution contributes to the creation of polymorphic connected ecosystems, evolving according to users' configurations. This problematic raises several questions in the field of modern forensics concerning the sharing and cross-referencing of useful information in order to accurately reconstruct the chronology of events in its context. It calls into question a static and unitary approach of the crime scene, by being part of a more global approach of the thing. The added value of the IoT comes from the fact that the whole is greater than the sum of its parts, which is why unitary approaches miss important information.

III. Analysis of a Crime Scene with IoT Devices

This section describes the process of extracting and analyzing data from multiple families of IoT devices at a crime scene in the manner of a classical approach in forensic. To illustrate our approach, we propose a scenario based on the real facts of the discovery of a dead body in an apartment.

A. Presentation of the Crime Scene

On 10 April 2018 at 8 a.m., police were alerted to a burglary and gun sounds coming from an apartment. A patrol arrived on the scene at 8:15 a.m. They discovered that the front door of the apartment had been forced open. The place also shows many signs of a struggle and violence. During the reconnaissance of the premises, the body of a lifeless person was found lying on a bed. The investigators therefore implemented the first protective measures by freezing the crime scene. A forensic team, including a computer scientist, takes over the crime scene at 9 a.m.

The apartment covers 45 m². It includes three separate rooms: an entrance (room 1), a bedroom (room 2) and a living room (room 3) (Fig. 1). It contains many connected objects belonging to several IoT architectures. It has a home automation system from an Orvibo kit. It contains two opening sensors (1 and 2); and a motion sensor (3) coupled to a Wi-Fi camera (4). This kit is located in the room 1 and on two exterior openings. This infrastructure communicates via ZigBee to a dedicated hub (5), located in the room 3. The home automation system is also made of Philips brand connected bulbs (6 and 7) with its dedicated hub (8). They are located in rooms 2 and 3 of the apartment. Four Sen.se Cookies are hidden in the different rooms. They transform household objects into connected objects. In our case, Cookies follow the water supply level of the coffee machine (9), the ambient temperature (10), the position of the bicycle (11) and the physical activity of the victim (12). All these objects are connected with a proprietary protocol to Mother Sen.se (13), in room 3. These different hubs, an Amazon Echo (14), a RaspberryPi0 (15) and an IP camera M136W (16) are

connected to the Internet by WinkHub 2 (17). The victim is lying on the bed in room 2. She has an Apple Watch series 3 (18) on her right arm and an iPhone SE (19) in her pocket. Hidden in the bed, there is a sleep sensor named Terraillon Dot (20). The apartment contains other objects such as Sens'it (21), a Heroz bracelet (22) and a Nokia Wi-Fi scale (23).

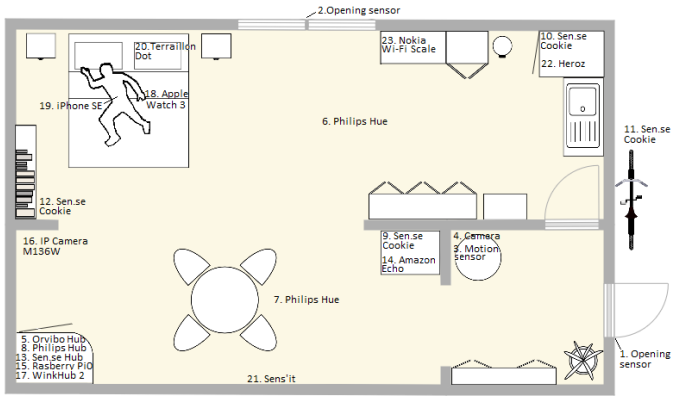

Fig. 1. Home layout with the IoT devices

B. Analysis Strategy

The study of IoT requires a multi-faceted approach to gather evidence from a variety of sources. They fall into three main groups [29]. Data from smart devices and sensors present at the crime scene make up the first group (SmartWatch, wellness and health devices, home automation, environmental measuring devices, etc.). The second digital environment includes information from hardware and software enabling communication between connected devices and the outside world (computers, mobile devices, firewalls, ...). The third group includes external resources from cloud platforms, social networks, Internet service providers and mobile networks.

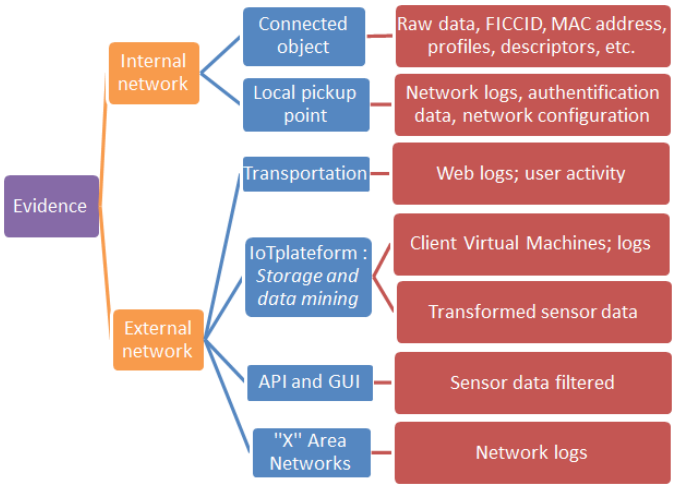

Fig. 2. Potential Sources of Evidence

The plurality of the IoT ecosystem offers a variety of data, rich in investigation: event logs, system-related information, network communication and telephony data, location, multimedia content, web history, environmental or activity measurements, deleted information, etc. (Fig. 2) [30]. However, the data are more or less directly accessible to police investigators. Indeed, data in the cloud requires a precise knowledge of the objects present locally in order to make a targeted request to online platforms. Thus, the first stage of forensic investigation focuses on groups 1 and 2. In addition, not all groups 1 and 2 devices are relevant to the study. It is necessary to prioritize the analysis of devices in relation to the elements sought for the study. Relevant data is identified, collected and retained without compromising its integrity. The heterogeneous nature of IoT devices makes the identification of data sources a difficult task, unlike traditional devices such as computers, servers or networks that contain some kind of storage media such as hard disks or USB sticks. They have their own data formats, protocols and physical interfaces [12]. In many cases the data is not stored on the device but on a connected service which may be a cloud-based system or on the same local network [28]. Some objects use automatic synchronization of their data with the network. As a result, they store locally little useful information for surveys. This is the case of the objects in the Philips, Orvibo and Sen.se kits. Relevant data is contained in the gateways. They are related to network activity and system configurations. The Orvibo camera is treated as an independent object on the network. In effect, it has an external storage space, which serves as a buffer when transmitting information to the network. Other objects are more versatile in their operation. These include the Apple Watch 3 and the Amazon Echo. Despite automatic synchronization of the object with the smartphone or the network, they still have locally relevant data. For the Terraillon Dot, a manual synchronization action must be performed through the user's application to bring up the data. Thus, the information can be redundant between media as long as there has been no memory rewriting. The smartphone concentrates a lot of relevant data by acting as a gateway and a user interface through the different object management applications. Once the data sources have been identified, the type of acquisition is determined, which is normally physical, logical or live forensics.

C. Local Data Extraction

The extraction of data from local devices is based on laboratory knowledge and techniques developed for forensic analysis of mobile equipment and embedded electronic systems. There are several levels of extraction providing access to different and complementary information: manual, logical and physical [31] (Fig. 3). Logical Extraction [32] tries to find the visible elements of the file system. This operation requires prior knowledge of the technical specifications of the devices and the version of the operating system (OS). This information directly affects the choice of the communication strategy to be adopted for performing the logical extraction: a connection to the Universal Serial Bus (USB), the use of serial or wireless protocols, an Application Programming Interface (API), proprietary commands, etc. This extraction relies on

the logical communication protocols between the target device and the analysis space. Thus, through the API, the investigator interacts directly with the device's OS. However, he can only request an extraction of data accessible only by the OS. In some cases, the device is placed in "diagnostic mode". For example, some versions of the Amazon Echo allow access to this mode by activating the ADB. This allows direct access to the system. Data is then retrieved using the manufacturer's protocols. The physical extraction is done at three abstraction levels: JTAG, chip-off[2] followed by the reading of the internal memory and FIB-SEM[3] micro-reading. It consists of a very low level copy of all the binary data physically present in the silicon of the equipment's memory. It results in an electronic reading of the state of all the elementary memory cells. Thus, it provides a collection of all the information still physically present in the flash memory. This is the crucial difference with logical extraction, which focuses on the allocated space.

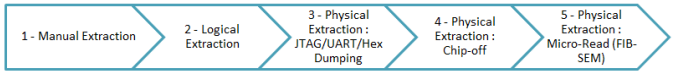

Fig. 3. Extraction levels organized by the level of difficulty

In the case of our crime scene, we proceeded with the logical and physical extraction of the data present in the objects (Tab. I). The choice of extraction type is determined by the information sought by the investigator, its technical feasibility, the time available and the cost of the intervention. By increasing the level of extraction, the cost of the tools and the operation increases. In addition, invasive and destructive techniques require a high level of expertise.

No.	Equipment	Manufacturer	Acquired data/image
4	Wi-Fi Camera	Orvibo	Physical extraction (SD card)
5	Hub	Orvibo	Logical extraction (Telnet)
8	Hub	Philips	Logical extraction (API)
13	Mother	Sen.se	Logical extraction (API)
14	Echo Spot	Amazon	Physical (JTAG and Chip-off) and Logical extractions (ADB)
15	Pi0	Rasberry	Physical extraction (SD card)
17	Hub 2	Wink	Logical extraction (API)
18	Watch 3	Apple	Logical extraction (Backup)
19	iPhone SE	Apple	Logical extraction

TABLE I
EXTRACTION LEVEL CHOSEN FOR THE DEVICES

The memories of the connected objects Sen.se, Orvibo, Philips, Sens'it, the Heroz wristband and the IP Camera M136W were not processed due to a small internal memory, an automatic data synchronization policy and their relevance to the investigation (proximity to the wanted event and position on the crime scene). Their identification elements are nevertheless relevant in the context of requisitioning from platform operators and analysis of events.

[2]Removing and reading the device's memory chip to read data and conduct analysis.

[3]Using high-power microscope to have a physical view of the device's memory cells to extract data.

The automation of decoding and the formatting of extraction results are facilitated by the use of classic digital forensics tools such as *X-Ways*, *Forensic Explorer*, *OSForensics*, *UFED*, etc. However, their uses require data or images in a standard or recognized format. The lack of a standard file structure makes the review process more difficult and time consuming. It requires data and file profiling. It is often accompanied by reverse engineering, especially when dealing with data compression or encryption mechanisms.

D. Type of Data Found in the Connected Equipment

The plurality of the connected ecosystem offers a variety of data produced and exchanged. This source of information can be broken down into functional data useful for rendering the service (multimedia content, telephony and location data, web history, environmental and activity measurements) and peripheral data related to the operation of the system and the network (event logs). Local devices also contain context data such as a physical and logical configuration of a place, a life habit, a sound or video recording of a specific phenomenon. These elements are combined with personal data such as the identity of the service consumer, his digital and biometric profile. This information allows investigators to qualify a phenomenon and to faithfully reconstruct the succession of events in its environment.

The home automation gateways *Orvibo* and *Philips* contain information on the user, on the ZigBee network and its addressing, on the physical and logical configuration of the home with its floors and rooms, on families of connected objects and their associations with the environment, on programed uses and security scenarios. They record all interactions with the infrastructure, categorized according to the type of action triggered: a command generated by a mobile application or by local objects, but also scenarios programed in response to a phenomenon observed. All these transcripts are time-stamped and filled in. From a forensic point of view, this information provides the investigator with information on when and how a presence or activity detected by the ecosystem and on the domestic use of living objects. It obtains an initial identification of the actors of the specific phenomenon and characterizes its trigger. The mobile application enriches the investigation of data registered in time in particular to understand the context but also on the situation of the phenomenon with the dates of the interactions and the material association. The external storage card of the camera *Orvibo* completes the reconciliation by a multimedia return.

The *Cookie* sensors measure changes in state such as movement or temperature. The *Mother* gateway contains event logs tracing communications and information back to the connected infrastructure. The application *Sen.se* associates a measurement of phenomena, with predetermined and configured information. It gathers data on the user, on the proprietary network and on the association logic between an object and an action via graphical renderings. The investigators obtain from this ecosystem a chronology of phenomena that occurred in

coherence with the use of a physical object, digitized by the *Cookie*.

In our use case, the *Amazon Echo* has the role of intelligent interface delivering commands to the infrastructure. The user can activate the connected light bulbs with a voice command. This interface mainly contains information about the user, the network and a set of activity logs related to the system, current operation, events and exchanges with the network. The mobile application concentrates the history of interactions and requests from the voice assistant. Thus, the event logs notify the action of capturing a sound at the time of the event. However, the sound recordings are stored on the *Amazon* cloud. Only their access links are filled in the application.

The analysis of the *WinkHub* is relevant to understand the architecture of the local network. This gateway contains all the network events and the identifiers of the connected equipment. It identifies and maps the path and the ascent of the data through the connected infrastructure to the Internet. The analysis of the dependency link with the *iPhone* is appropriate for digital investigations. This connection informs geographical proximity of the smartphone.

The *Terraillon Dot* records the movements on a flat surface, here on the bed. It contains the latest measurements taken, network configuration and synchronization information. The uploading of data is triggered manually from the mobile application. It contains data on the user (numerical user profile and health information), the network, synchronization events and all health measurements over the last 30 days (sleep duration, lift/bed rest, body movement, etc.). The *Apple Watch* complements and solidifies these health measurements in connection with physical activity and location elements. It gives an accurate picture of habits and contextualized events. It also gathers information on the user, the network, telephony, messaging and Internet browsing.

The different connected ecosystems provide a wealth of information, a phenomenon of digitization of a user's profile in time and space with its precise context. The redundancy and cross-referencing of data in the different parts of the infrastructure offers a certain fidelity, reliability and weighting of the configured elements. This information must be corroborated with the elements present on the Cloud platforms (Timeline logs, port scans, metadata logs, control node logs, interface logs, Runtime logs, etc.). They can bring missing or complementary information in the understanding of an event. These spaces also convey novel associations on the user's digital profile. For example, the *Amazon* platform contains data about the connected house, but also links to other objects attached to the user account, its browsing and purchasing habits, its uses (e.g. reminders or alarms), its personal directories (drive), location, etc. The only constraint in the processing of this source of information lies in the legal limits.

Once evidence is successfully collected and decoded from IoT devices and within its infrastructure, regardless of the file system, OS or platform on which it is based, the work of concentrating and contextualizing the data is necessary. The investigator tries to understand the path the data has taken from its creation to its distribution within the infrastructure and the reasons for its position during analysis. The objective of this approach is to qualify for the data and to establish the coherence of the traces captured with regard to the criminal act.

IV. Contextualization and Valuation of Data

The traces obtained are made up of numerous fragments contained in a plurality of supports of the same network, likely to evolve in time and space. In order not to be fragmented, the analysis must not focus on the study of a single object but on the ecosystem as a whole. It thus consists of studying and thinking the data according to three axes: the **time** by defining the chronology of the events and the iterative phenomena, the **space** by positioning the data in the infrastructure and by taking into account the local environment and the **context** by analyzing the event with regard to the roles and the actions of the various equipment. Through this ternary approach, the investigator seeks to determine the lifecycle of the data in order to qualify them and establish their consistency. We assume that there is a causal dependency between the different events and the state of the data in the system.

In order to perform a meaningful analysis of the events, it is necessary to observe the information according to a common timestamp. The investigator must make sure to find the timestamp of each piece of equipment or the method of synchronization of the systems. He calculates the difference between this hardware data and the world clock in order to integrate this differential in his reasoning.

A. Hidden Links and Dependencies

In a connected and synchronized ecosystem, an event is the result of several interactions between digital devices. The information generated is disseminated to a network of interdependent devices. Thus, this data is potentially stored on one or more media located outside the primary ecosystem. By working in a unitary way on each media, the correlation between the devices is lost. In order to understand this issue, the investigator must integrate in his analysis phase the study of the network architecture (*IoT network monitoring*) by identifying the dependencies and the roles of the equipment. He must also model the activity based on the study of the systems' event logs and interactions.

1) Identification of Connected Equipment: In a generic way, the local environment is composed of connected objects and gateways (Tab. II). Connected objects are more or less elaborate. Some behave as simple **sensors** or **actors** by measuring, detecting and reacting to some data or commands of the physical environment. Other **objects** have more computing and storage capacity with more or less autonomy to interact with the network, as for example IP cameras or smart TVs. In addition to these functional characteristics, there is a factor of dependency of objects to communicate. Objects can be distinguished by their ability to exchange data directly outside their ecosystems or through gateways. Gateways can also be classified according to their roles in the infrastructure: either as

a ”network node” or as an ”interface’ with the outside world. Nevertheless, these distinctions are more and more complex and put at fault by the development of hybrid solutions, in particular with the decentralization of the treatment by the phenomena of *edge computing* for the gateways and *fog computing* for connected objects. Moreover, the local ecosystem is governed by a **controller**. It is potentially an integral part of a gateway, as in the case of connected stations. It can be a separate device, for example in the form of a mobile application interface. The Internet of things is completed by a **cloud service** real crossroads of information.

Connected object	Sensor or actuators	Sensors Sen.se (cookie), Philips light bulbs, Orvibo opening and presence sensors Orvibo camera, IP Camera M136W, Terraillon Dot, Sens’it, Heroz and Nokia scale
	Advanced object	Amazon Echo, Raspberry Pi0 and Apple Watch
Gateway	Node Interface to the Internet	Sen.se Mother, Orvibo and Philips WinkHub2 and iPhone
HMI	Controller	Amazon Echo and apps (iPhone)

TABLE II
GENERIC CLASSIFICATION OF CRIME SCENE EQUIPMENT

2) Network events and hidden links: The network activity between objects, nodes and gateways is accompanied by three main families of logs: equipment status updates, action commands and information about the network operation. The equipment status update contains the state of the sensors at a given time. This information is sent by the objects to the gateways sequentially and at regular intervals. The action command log contains the commands sent by the users.The frames consist of at least a sender, a receiver and a description of the command. By parsing the payload attribute and embedded log values, the action command of the devices is identified and extracted [33]. The log associated with the network activity contains information about the network status. It traces pairings and all exchanges between devices on the same network. It is summarized in the form of frames containing the device identification and a network activity value. The gateways form a two-way communication that acts as a centralizing messenger linking the various media and cloud platforms. By recording network activity passing in both directions, events are identifiable and extracted.

The study of hidden links between devices is based on the characterization of multiple events sharing a single attribute. For example, by studying the event logs contained in the gateway and the application *Philips*, the event ”Light on” is time stamped. However, it must be characterized more precisely. Is it a user action through the phone application, an external switch, a sensor signal or a voice command transmitted by the *Amazon Echo*? Was it performed by a known user? Is it a programed action? For each question, a unique data is associated. It characterizes the event whether it is due to an active or passive human interaction, without interference, direct or not. This analytical approach can be generalized to all connected objects in the infrastructure. It allows us to determine hidden links, here an association between the *Amazon Echo* and the *Philips* solution following a voice command *Alexa*. To do this, the event is represented using a graphical model (Fig. 4). This approach simplifies the correlation process and the grouping of events. By combining the timestamp, the groups of the same temporal dimension are gathered and then compared using the common attributes. In a more global way, we obtain a new modeling of the network traffic of the crime scene based on the orientation of the network and its attributes (Fig. 5). This graphical representation models the flow of communication messages and helps identify the sources of evidence (Tab. III). The investigator determines the actors of the event, their positions, the actions performed or detected and the response of the objects to the different solicitations. This unique attribute is declined according to the identifiers of a user, an object, a place or an exchanged data. The investigator will try to link the different connected equipment to this common parameter according to a heuristic approach. This approach seeks to discriminate irrelevant elements and to adjust the analysis.

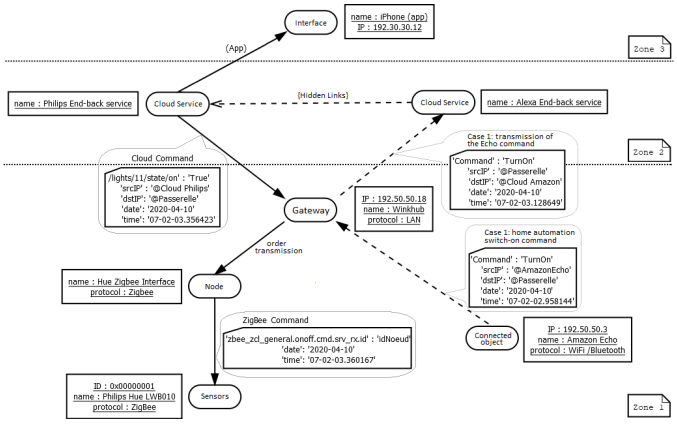

Fig. 4. Modeling of the ‘light on’ action following a voice command *Alexa*

Source	Description	Path
Gateway Philips	- Cloud connection status - Date of the first and last interaction performed by the user - List of people authorized to interact	`/config/`
	- Programed or configured events	`/schedules/` and `/scenes/`
Amazon Echo	- User account information (name and type of service)	`/system/users/ : 0.xml` and `/0/accounts.db (P16)`
	- Activity logs	`/system/dropbox/`
Philips (App.)	- Recording of actions performed by the user from the application	`com.philips.lighting.hue2 : /Library/com.amplitude.database/`
	- Activity logs	`group.com.philips.hue2 : /debuglog/`
Amazon Echo (App.)	- History of the discussions	`AlexaMobileiOSComms.sqlite`
	- Map log and voice interaction history	`RCTAsyncLocalStorage_V1`

TABLE III
”LIGHT ON” EVENT MARKERS.

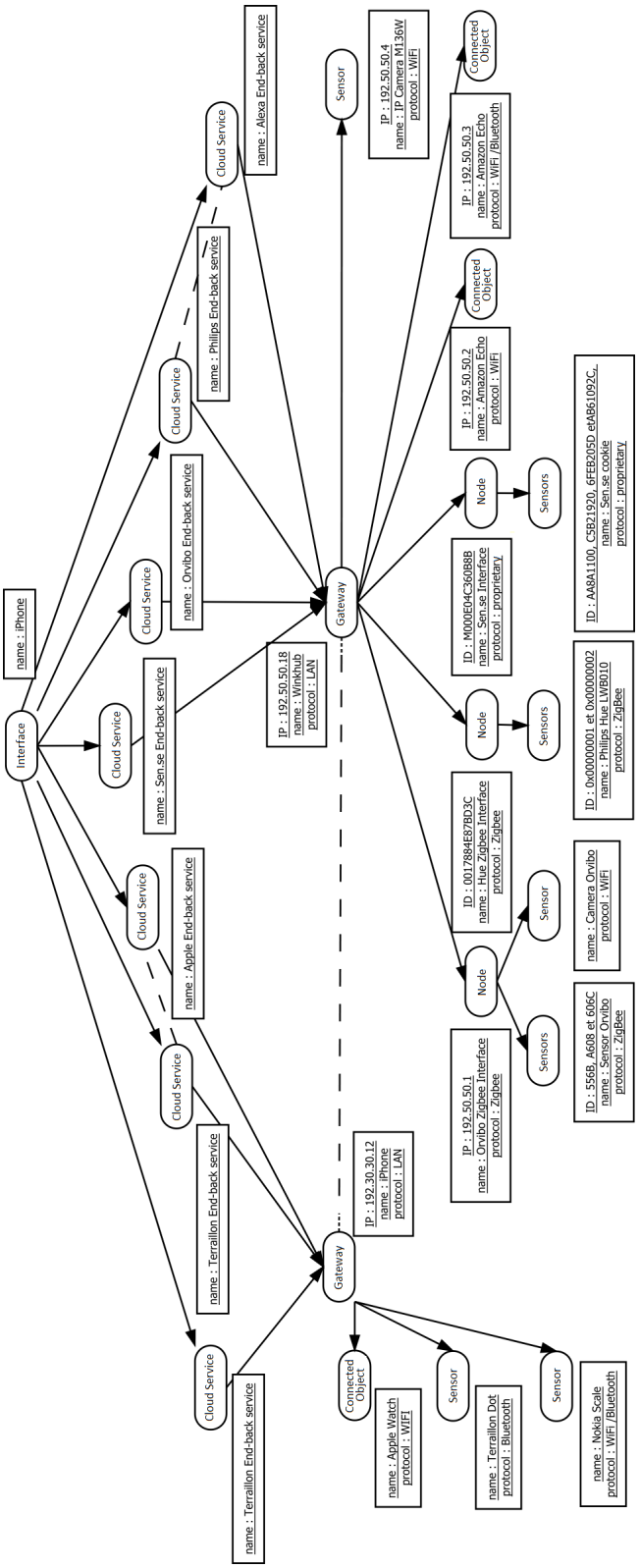

Fig. 5. General mapping of the connected environment

B. Chronology of the Criminal Phenomenon

From the data collected, the investigator is able to date the intrusion of the victim's home with the *Orvibo* sensor of the entrance gate. He materializes the travel of the respondent with the home automation system *Orvibo*, corroborated by the movement of the sensor *Sen.se Cookie*. It determines the time of death of the victim from the health data of the connected watch (Tab. IV). He notes the absence of modification of the crime scene in particular in the post-mortem movements of the body by crossing the data of the *Terraillon Dot* and the *Apple Watch*. With all these elements, he is able to emit hypotheses in the place of the murder and the circumstances. This information must be cross-referenced with forensic data collected from the crime scene and the victim.

Date	Description	Source	Events
T0 10/04/20	Lights off (6-7) / Door and window closed (1-2) / Health measure: cardiac activity (18) and movement (20)	Philips and Orvibo Hubs / iPhone (Health - Home Automation) / Terraillon Dot	One person in room 2 (known)
06:43:17	Door opening detected (1)	Orvibo Hub / iPhone (Home Automation) / WinkHub	Presence of people in room 2 (known) and room 1 (not recognized). Movements in room 1
06:44:03	Motion detected (3)		
06:44:12	Camera activated (4)	Orvibo Camera / WinkHub / Orvibo Hub / iPhone (Home Automation)	
06:52:46	Cookie activated (9)	Mother / WinkHub / iPhone (Sen.se)	Presence of people in room 2 (known) and room 3 (not recognized). Movements in room 3
06:54:16	Camera activated (6)	WinkHub / iPhone (IP Camera)	
06:57:11	Movement (20)	Terraillon Dot	
07:01:04	Heart rate acceleration (18)	iPhone (Health)	
07:02:02	Voice control (14)	Amazon Echo / WinkHub	
07:02:03	Light on (6)	WinkHub / Philips Hub / iPhone (Home Automation)	
07:07:01	Cardiac arrest (8)	iPhone (Health)	Presence of one person in room 2 (known).
07:07:54	End of the movement (20)	Terraillon Dot	
07:11:44	Motion detected (3)	Orvibo Hub / iPhone (Home Automation) / WinkHub	Presence of people in room 2 (known) and room 3 then 1 (not recognized). Movements in rooms 3 and 1 (3)
08:17:21	Motion detected: arrival of the patrol (3)		
08:24:56	Motion detected (3)		
09:12:00	Detection of movement: arrival of the forensic team (3)		
T1	Lights: on (6) off (7) / Door open (1) / Windows closed (2)	Philips and Orvibo Hubs / WinkHub / iPhone (Health - Home Automation)	

TABLE IV
CHRONOLOGY OF EVENTS MEASURED BY CRIME SCENE IOT DEVICES

In addition, digital data can be used to guide certain investigations, such as the collection of papillary and biological traces at the crime scene, based on the criminal history of the accused. On a larger scale, this information can be used to delimit and discriminate a study area by defining an investigation strategy. Thus, connected objects make it possible to verify working hypotheses by providing new material elements, such as a motive that is incompatible with an operating mode. The study of the chronology of events can also provide information on a possible premeditation in criminal logic.

V. CONCLUSION

The Internet of Things constitutes an unprecended receptacle of information and data. This phenomenon generates signficant challenges and opportunities for digital investigation.

Many approaches consider only the object and try to move traditional digital forensic to connected devices. However, for us, the main challenge is to find consistency in scattered data and hidden links between heterogeneous objets.

In this paper, we present our methodology to find "hidden link" between objects, action and data. By this means, we can discover more valuable information to understand criminal act. We based our approach on real case and prove that the correlation between data can produce new and more information. To improve our methodology, we currently implement new mobile tools usable directly on the scene crime. In order to respond with relevance to the needs of investigation and to understand past phenomena, the technician in new technologies must look at the data coming from multi-sources according to the prism of time, space and context. He relies on the study of the network architecture, on the role of each particular equipment in the actions taken and on the understanding of the data migration within the infrastructure. The cross-referencing of intelligently obtained traces guarantees unprecedented checks and investigations to identify people, places, events and elements associated with the legal case.

To enhance our solution, we have also started to work on the representation of the different events and their federation based on time stamps. From this, we conduct research on automatic link discovery based on graph theory and machine learning.

REFERENCES

[1] B. Dorsemaine, J.-P. Gaulier, J.-P. Wary, N. Kheir, and P. Urien, "Internet of things: a definition & taxonomy," in 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies. IEEE, 2015, pp. 72–77.

[2] V. Katalov and M. Epifani, "Apple watch forensics: Is it ever possible, and what is the profit?" https://www.forensicfocus.com/news/apple-watch-forensics-is-it-ever-possible-and-what-is-the-profit-2/, 2019.

[3] D. H. Kasukurti and S. Patil, "Wearable device forensic: Probable case studies and proposed methodology," in International Symposium on Security in Computing and Communication. Springer, 2018, pp. 290–300.

[4] S. Kang, S. Kim, and J. Kim, "Forensic analysis for iot fitness trackers and its application," Peer-to-Peer Networking and Applications, vol. 13, no. 2, pp. 564–573, 2020.

[5] Á. MacDermott, S. Lea, F. Iqbal, I. Idowu, and B. Shah, "Forensic analysis of wearable devices: Fitbit, garmin and hetp watches," in 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE, 2019, pp. 1–6.

[6] H. Chung, J. Park, and S. Lee, "Digital forensic approaches for amazon alexa ecosystem," Digital Investigation, vol. 22, pp. S15–S25, 2017.

[7] S. Li, K.-K. R. Choo, Q. Sun, W. J. Buchanan, and J. Cao, "Iot forensics: Amazon echo as a use case," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6487–6497, 2019.

[8] S. Tristan, S. Sharma, and R. Gonzalez, "Alexa/google home forensics," in Digital Forensic Education. Springer, 2020, pp. 101–121.

[9] P. van Bolhuis and C. Van Bockhaven, "Forensic analysis of chromecast and miracast devices," Cybercrime and Forensics Project, Master's Program in System and Network Engineering, University of Amsterdam, Amsterdam, The Netherlands, 2014.

[10] N. K. Bharadwaj and U. Singh, "Acquisition and analysis of forensic artifacts from raspberry pi an internet of things prototype platform," in Recent Findings in Intelligent Computing Techniques. Springer, 2019, pp. 311–322.

[11] T. Zia, P. Liu, and W. Han, "Application-specific digital forensics investigative model in internet of things (iot)," in Proceedings of the 12th International Conference on Availability, Reliability and Security, 2017, pp. 1–7.

[12] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," Ad hoc networks, vol. 10, no. 7, pp. 1497–1516, 2012.

[13] J. T. Rajewski, "Internet of things forensics," A Presentation at Enfuse, 2016.

[14] J. Boucher and N.-A. Le-Khac, "Forensic framework to identify local vs synced artefacts," Digital Investigation, vol. 24, pp. S68–S75, 2018.

[15] A. Goudbeek, K.-K. R. Choo, and N.-A. Le-Khac, "A forensic investigation framework for smart home environment," in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2018, pp. 1446–1451.

[16] X. Zhang, K.-K. R. Choo, and N. L. Beebe, "How do i share my iot forensic experience with the broader community? an automated knowledge sharing iot forensic platform," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6850–6861, 2019.

[17] Y. Amar, H. Haddadi, R. Mortier, A. Brown, J. Colley, and A. Crabtree, "An analysis of home iot network traffic and behaviour," arXiv preprint arXiv:1803.05368, 2018.

[18] B. Copos, K. Levitt, M. Bishop, and J. Rowe, "Is anybody home? inferring activity from smart home network traffic," in 2016 IEEE Security and Privacy Workshops (SPW). IEEE, 2016, pp. 245–251.

[19] M. R. Santos, R. M. Andrade, D. G. Gomes, and A. C. Callado, "An efficient approach for device identification and traffic classification in iot ecosystems," in 2018 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2018, pp. 00 304–00 309.

[20] J. He, C. Chang, P. He, and M. S. Pathan, "Network forensics method based on evidence graph and vulnerability reasoning," Future Internet, vol. 8, no. 4, p. 54, 2016.

[21] P. Neise, "Graph-based event correlation for network security defense," Ph.D. dissertation, The George Washington University, 2018.

[22] D. Barrett and G. Kipper, Virtualization and forensics: A digital forensic investigator's guide to virtual environments. Syngress, 2010.

[23] C. Esposito, A. Castiglione, F. Pop, and K.-K. R. Choo, "Challenges of connecting edge and cloud computing: A security and forensic perspective," IEEE Cloud Computing, vol. 4, no. 2, pp. 13–17, 2017.

[24] F. Zafar, A. Khan, S. U. R. Malik, M. Ahmed, A. Anjum, M. I. Khan, N. Javed, M. Alam, and F. Jamil, "A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends," Computers & Security, vol. 65, pp. 29–49, 2017.

[25] E. Oriwoh, D. Jazani, G. Epiphaniou, and P. Sant, "Internet of things forensics: Challenges and approaches," in 9th IEEE International Conference on Collaborative computing: networking, Applications and Worksharing. IEEE, 2013, pp. 608–615.

[26] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," IEEE transactions on parallel and distributed systems, vol. 22, no. 5, pp. 847–859, 2010.

[27] Q. Alam, S. U. Malik, A. Akhunzada, K.-K. R. Choo, S. Tabbasum, and M. Alam, "A cross tenant access control (ctac) model for cloud computing: Formal specification and verification," IEEE Transactions on Information Forensics and Security, vol. 12, no. 6, pp. 1259–1268, 2016.

[28] A. Attwood, M. Merabti, P. Fergus, and O. Abuelmaatti, "Sccir: Smart cities critical infrastructure response framework," in 2011 Developments in E-systems Engineering. IEEE, 2011, pp. 460–464.

[29] S. Zawoad and R. Hasan, "Faiot: Towards building a forensics aware eco system for the internet of things," in 2015 IEEE International Conference on Services Computing. IEEE, 2015, pp. 279–284.

[30] F. Bouchaud, G. Grimaud, T. Vantroys, and P. Buret, "Digital investigation of iot devices in the criminal scene," Journal of Universal Computer Science, vol. 25, no. 9, pp. 1199–1218, sep 2019, http://www.jucs.org/jucs_25_9/digital_investigation_of_iot.

[31] S. Brothers, "How cell phone" forensic" tools actually work-proposed leveling system," in Mobile Forensics World Conference, Chicago, Illinois, 2009.

[32] K. Kim, D. Hong, K. Chung, and J.-C. Ryou, "Data acquisition from cell phone using logical approach," Proceedings of the world academy of science, engineering and technology, vol. 26, 2007.

[33] Y. Jia, Y. Xiao, J. Yu, X. Cheng, Z. Liang, and Z. Wan, "A novel graph-based mechanism for identifying traffic vulnerabilities in smart home iot," in IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018, pp. 1493–1501.